D8A CHEATSHEET · #045 · DISTRIBUTIONS

Six shapes explain most of your data.

D8A.ACADEMY/RESOURCES/PROBABILITY-DISTRIBUTIONS

NORMAL

±1σ ≈ 68%	±2σ ≈ 95%
heights, errors	sums of small causes
CLT	means → normal, n≥30

BINOMIAL

n trials, p each	count of successes
mean = np	var = np(1-p)
conversions, churn	yes/no repeated

POISSON

events per window	rate λ
mean = var = λ	the tell
tickets, arrivals	rare + independent

LONG TAILS

log-normal	salaries, revenue
mean >> median	the skew tell
log(x) first	then it behaves

TWO MORE

uniform	anything equally likely
exponential	time between events
memoryless	its famous property

GOTCHAS

assume normal	← plot it first.
mean of a long tail	one whale owns it
outliers at ±3σ	only if truly normal

WORTH MEMORIZING

01 Match the data to the shape

The interview version: name the variable, name the distribution, name why.

conversion rate	Binomial: n visitors, each converts with probability p.
tickets per hour	Poisson: rare independent events in a fixed window.
revenue per user	Log-normal: most pay little, a few whales pay a lot.
time between fails	Exponential: the waiting-time twin of Poisson.
mean of samples	Normal, by the CLT, almost whatever the source shape.

02 Trap: treating everything as normal

The 68/95 shortcuts and z-score fences are only true for the bell shape.

YOU ASSUMED	REALITY	DO INSTEAD
revenue is normal	log-normal: whales rule the mean	median, or log(x) first
±3σ finds outliers	long tails live past 3σ legally	plot, then pick fences
mean = typical user	mean can beat the 90th pctile	report percentiles

→ One histogram costs ten seconds and prevents every mistake in this table.